Docket No.: 0112855.00122US2
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant:	Jordan COHEN et al.	Confirmation No.:	9023
Application No.:	10/657,421	Art Unit:	2626
Filed:	September 8, 2003	Examiner:	P. D. Shah
Title:	PROSODIC MIMIC METHOD AND APPARATUS		

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

APPEAL BRIEF

Dear Sir:

As required under § 41.37(a), this brief is filed after the Notice of Appeal filed in this case on March 29, 2010, and is in furtherance of said Notice of Appeal.

The fees required under § 41.20(b)(2), and any required petition for extension of time for filing this brief and fees therefor, are dealt with in the accompanying TRANSMITTAL OF APPEAL BRIEF.

This brief contains items under the following headings as required by 37 C.F.R. § 41.37 and M.P.E.P. § 1205.2:

I.	Real Party In Interest
II	Related Appeals and Interferences
III.	Status of Claims
IV.	Status of Amendments
V.	Summary of Claimed Subject Matter
VI.	Grounds of Rejection to be Reviewed on Appeal
VII.	Argument

VIII. Claims
Appendix A Claims
Appendix B Evidence
Appendix C Related Proceedings

I. REAL PARTY IN INTEREST

The real party in interest for this appeal is:

Voice Signal Technologies, Inc., which is a wholly-owned subsidiary of Nuance Communications, Inc., 1 Wayside Road, Burlington, MA 01803.

II. RELATED APPEALS AND INTERFERENCES

There are no other appeals, interferences, or judicial proceedings which will directly affect or be directly affected by or have a bearing on the Board's decision in this appeal.

III. STATUS OF CLAIMS

A. Total Number of Claims in Application

There are 11 claims pending in application.

B. Current Status of Claims

1. Claims canceled: 5, 10, 13, and 14

2. Claims withdrawn from consideration but not canceled: None

3. Claims pending: 1-4, 6-9, 11-12, and 15

4. Claims allowed: None

5. Claims rejected: 1-4, 6-9, 11-12 and 15

C. Claims On Appeal

The claims on appeal are claims 1-4, 6-9, 11-12 and 15

IV. STATUS OF AMENDMENTS

All amendments that have been submitted have been accepted and Appendix A presents the pending claims including all amendments that have been accepted.

V. SUMMARY OF CLAIMED SUBJECT MATTER

The invention relates to a telephone device and implemented method which extracts and uses prosodic features of a user's spoken words to synthesize and generate an audible output that is high quality, realistic-sounding speech that sounds like the user's voice. One specific application involves improving the quality and intelligibility of synthesized voice messages used to confirm spoken commands of a mobile telephone user. (see ¶ [0018]).

Referring to Fig. 1, an input device, such as a microphone, captures a spoken utterance 102 (for example, the phrase "CALL HOME"). The spoken utterance 102 corresponds to an action to be taken by the mobile telephone device. In this example, the telephone looks up and dials the telephone number for (HOME). (see ¶ [0029]).

The system analyzes spoken utterance 102 for its prosodic parameters (e.g. pitch of the spoken utterance) and extracts values for the prosodic parameters. Using an MFCC analyzer, the system also extracts the spectral content, e.g., mel cepstra, and energy content of spoken utterance 102. The MFCC analyzer outputs frames of prosodic parameters. (see ¶'s [0030] and [0031]).

A decoder or speech recognition engine, employing hardware and software to select a recognized word from a set of possible known words, decodes or recognizes the spoken utterance. Upon recognizing the spoken word, the decoder provides the word as a text output 132 to a display device to visually indicate the results of the decoding. (see ¶ [0032]).

The decoder also delivers the recognized word 134 to a speech synthesizer that uses the recognized word and a set of default programmed (nominal) synthesis rules to generate synthesized nominal word frames. For example, the decoder might use a whole-word model, in which case the synthesis takes place at the word level. (see ¶ [0033]).

Using the recognized word's nominal synthesized frames 142, the captured prosodic parameters provided in the pitch per frame 112 and the actual frames 124, a prosodic mimic generator generates the prosodic mimic phrase. The prosodic mimic generator applies the prosodic parameters to the nominal frames 142 on a frame-by-frame basis and also temporally aligns the generated mimic word with the nominal word at a whole-word level. In other words, in one particular embodiment, the recognized word 134 is aligned in time with the corresponding captured spoken word by forcing the start and end points of the nominal word to correspond to those of the spoken word. The result is a synthesized prosodic mimic phrase that, owing to its prosody, mimics the original spoken word in its content and its sound. (see ¶'s [0034] and [0035]).

An audio converter receives the generated prosodic mimic phrase and converts the nominal frames with the applied actual timing and pitch 152 into an audio signal to be played on the mobile telephone's speaker, which is the same speaker over which the user hears the ordinary telephone communication output. (see ¶ [0036]).

The end result of the process described above is a natural-sounding audible phrase resembling the originally spoken utterance 102. This synthesized mimic phrase is used as an audible confirmation message played back to the mobile telephone user to confirm the command to be carried out or the name to be dialed. (see ¶ [0037]).

Claims 1 and 9 are presented in the following tables which map the recited elements to the relevant portions of the specification and figures:

Features of Claim	Support in Specification
1. A method for speech synthesis, said method implemented on a handheld device and comprising:	see process for synthesizing speech depicted in Figs. 2 and 3 see mobile telephone device 10 in Fig. 1
receiving a spoken utterance including at least one of a command to be executed by the handheld device and a name to be dialed by the handheld device;	see speech input 100 in Figs. 2 and 3 see page 8, lines 27-33
in response to receiving the spoken utterance:	Figs. 2 and 3
extracting one or more prosodic parameters from the spoken utterance;	see pitch detection 110 in Fig. 2 see page 4, lines 21-23 and page 6, lines 5-8
performing speech recognition on the spoken utterance to generate a recognized word;	see decoder/speech recognition engine 2300 in Fig. 1 see page 4, lines 27-29 and page 6, lines 12-16
from the recognized word that is generated from the speech recognition,	see speech synthesizer 2400 in Fig. 1

Features of Claim	Support in Specification
synthesizing a nominal word;	see page 6, lines 17-20
generating a prosodic mimic word from the synthesized nominal word and the extracted one or more prosodic parameters, wherein generating the prosodic mimic also involves temporally aligning the synthesized nominal word with the spoken utterance; and	see prosodic mimic generator 2600 in Fig. 1 see page 6, lines 21-28
if the recognized word includes the command, executing the command on the handheld device, and if the recognized word includes the name, dialing a number corresponding to the name.	see page 5, lines 31-34 and page 7, lines 9-12

Features of Claim	Support in Specification
9. A handheld system for speech synthesis, said system comprising:	see mobile telephone device 10 in Fig. 1 see process for synthesizing speech depicted in Figs. 2 and 3
an audio input device capable of receiving a spoken utterance including at least one of a command to be executed by the handheld system and a name to be dialed by the handheld system;	see audio input device 1000 in Fig. 1 see page 8, lines 27-33
a signal processor that, in response to	see pitch detector 2100 in Fig. 1

Features of Claim	Support in Specification
receiving the spoken utterance, determines one or more prosodic parameters of the spoken utterance;	see page 4, lines 21-23 and page 6, lines 5-8
a speech recognizer that, in response to receiving the spoken utterance, recognizes the spoken utterance and generates a corresponding recognized word;	see decoder/speech recognition engine 2300 in Fig. 1 see page 4, lines 27-29 and page 6, lines 12-16
a speech synthesizer that synthesizes a nominal word from the recognized word;	see speech synthesizer 2400 in Fig. 1 see page 6, lines 17-20
a prosodic mimic generator that receives the synthesized nominal word and the one or more prosodic parameters and generates a prosodic mimic word therefrom, said prosodic mimic generator also temporally aligning the prosodic mimic word with the spoken utterance, and	see prosodic mimic generator 2600 in Fig. 1 see page 6, lines 21-28
a processor that, if the recognized word includes the command, executes the command, and, if the recognized word includes the name to be dialed, dials a number corresponding to the name.	see processor 20 in Fig. 1 see page 5, lines 31-34 and page 7, lines 9-12

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

The examiner rejected claim 1-3, 6-7, 9, 12 and 15 under 35. U.S.C. § 103(a) as being unpatentable over European Patent Application EP 1,271,469 to Marasek et al., in view of U. S.

Patent No. 5,796,916 to Meredith (Meredith), in view of U.S. 6,081,780 to Lumelsky (Lumelsky), in view of International Publication No. WO 02/097590 to Cameron (Cameron).

VII. ARGUMENT

Claims 1 and 9:

For the reasons presented below, we submit that combing Marasek's system with the teachings of Meredith, Lumelsky, and Cameron, does not produce the invention of claims 1 and 9.

The examiner argues that Marasek performs speech recognition on the spoken word to generate a recognized word and in support of that he directs our attention to Fig. 1, step S12 and ¶ [0040]. Then, the examiner argues that in Marasek's system "speech synthesis is performed on the input speech by applying prosody to a given text" and in support of this the examiner appears to point to steps S40 and S50 in Fig. 1 as well as to [0046].

We note, however, that the examiner's characterization of the teachings of Marasek is not accurate. Marasek does not perform speech synthesis on input speech, he performs it when generating speech in, for example, a dialogue system (e.g. see [0024]), which involves producing speech from templates or stored responses. Marasek never even hints at performing speech synthesis to reproduce the very same speech (or word) that was received and in response to receiving that speech (or word). That would involve receiving an utterance and then playing that received utterance back as synthesized speech, something that Marasek does not do and is not described by Marasek.

In addition, applying prosody "to a given text" is not the same as applying prosody to a word that is synthesized from a recognized word. More specifically, it is not the same as "generating a prosodic mimic word from the synthesized nominal word and the extracted one or more prosodic parameters," wherein the synthesized nominal word is synthesized from the recognized word, as is required by the claims. Marasek makes clear that his system is for constructing personality patterns that can be used later when generating synthesized speech. There is no suggestion by Marasek that

it be used to synthesize the very same speech that was just received, which of course already has the personality pattern of the speaker. Marasek makes his approach clear in the following statement:

> The speech features are then directly or indirectly used to construct a personality pattern which can later on be used to reconstruct a speech output with the mimic of the speech input and its speaker. ¶ [0006] (emphasis added).

There is no need in Marasek to synthesize the vary same speech that was just received since it already most accurately reflects the sounds of the speaker, i.e., it already has the prosody of the speaker. Certainly, Marasek has not provided a reason for adopting such an approach.

It is worth reiterating a major difference between the claimed invention and the teachings of Marasek. Marasek says nothing about performing a number of steps in response to receiving a spoken utterance, including generating a recognized word from that spoken utterance and then synthesizing that recognized word to generate a synthesized word. More specifically, Marasek says nothing about performing at least two steps "in response to receiving the spoken utterance," wherein those steps involve: "performing speech recognition on the spoken utterance to generate a recognized word" and then "from the recognized word that is generated from the speech recognition, synthesizing a nominal word." Marasek simply says that the extracted prosody can be used later to synthesize words which have the personality of a particular speaker.

It is conceivable that Marasek would later synthesize a word that is the same as the recognized word in the received utterance. However, that is not what is claimed. That later synthesis of that word is not in response to receiving the original utterance containing that same word.

One aspect of the invention which is of particular value is performing both of those steps in response to receiving an utterance. This enables a mobile device to provide audio feedback to the user indicating whether the device has correctly recognized the utterance and to do so using prosody that most sounds like that of the speaker, a prosody that is more likely to be intelligible to the user. The specification explains that point this way:

[0011] These and other aspects of the invention provide improved speech synthesis, especially in small mobile devices such as mobile telephones with voice activated commands and user interfaces. In one respect, better synthesis of audible confirmation messages is enabled, the audible confirmation messages having prosodic attributes resembling those of the user. Better speech synthesis sounds more natural and is more understandable to humans, therefore the present invention improves the usefulness and intelligibility of audible user interfaces.

We note that claim 9 includes limitations similar to those discussed above in connection with claim 1 and thus similar reasoning applies.

For the reasons presented above, we submit that the claims are in condition for allowance and therefore ask that they be allowed to issue.

VIII. CLAIMS

A copy of the claims involved in the present appeal is attached hereto as Appendix A.

Applicant believes no fee is due with this response. However, if a fee is due, please charge our Deposit Account No. 08-0219, under Order No. 0112855.00122US2 from which the undersigned is authorized to draw.

Respectfully submitted,

Dated: September 22, 2010

/Eric L. Prahl/

Eric L. Prahl
Registration No.: 32,590
Attorney for Applicant(s)

Wilmer Cutler Pickering Hale and Dorr LLP
60 State Street
Boston, Massachusetts 02109
(617) 526-6000 (telephone)
(617) 526-5000 (facsimile)

CLAIMS - APPENDIX A

Claims Involved in the Appeal of Application Serial No. 10/657,421

1. (Previously Presented) A method for speech synthesis, said method implemented on a handheld device and comprising:

receiving a spoken utterance including at least one of a command to be executed by the handheld device and a name to be dialed by the handheld device;

in response to receiving the spoken utterance:

extracting one or more prosodic parameters from the spoken utterance;

performing speech recognition on the spoken utterance to generate a recognized word;

from the recognized word that is generated from the speech recognition, synthesizing a nominal word;

generating a prosodic mimic word from the synthesized nominal word and the extracted one or more prosodic parameters, wherein generating the prosodic mimic also involves temporally aligning the synthesized nominal word with the spoken utterance; and

if the recognized word includes the command, executing the command on the handheld device, and if the recognized word includes the name, dialing a number corresponding to the name.

2. (Original) The method of claim 1, wherein the one or more prosodic parameters include pitch.

3. (Original) The method of claim 1, wherein the one or more prosodic parameters include timing.

4. (Original) The method of claim 1, wherein the one or more prosodic parameters include energy.

5. (Canceled)

6. (Original) The method of claim 1, further comprising temporally aligning phones of the spoken utterance and phones of the nominal word.

7. (Original) The method of claim 1, further comprising converting the prosodic mimic word into a corresponding audio signal.

8. (Original) The method of claim 1, wherein the spoken utterance is received by a telephone input device and the prosodic mimic word is provided to a telephone output device.

9. (Previously Presented) A handheld system for speech synthesis, said system comprising:

an audio input device capable of receiving a spoken utterance including at least one of a command to be executed by the handheld system and a name to be dialed by the handheld system;

a signal processor that, in response to receiving the spoken utterance, determines one or more prosodic parameters of the spoken utterance;

a speech recognizer that, in response to receiving the spoken utterance, recognizes the spoken utterance and generates a corresponding recognized word;

a speech synthesizer that synthesizes a nominal word from the recognized word;

a prosodic mimic generator that receives the synthesized nominal word and the one or more prosodic parameters and generates a prosodic mimic word therefrom, said prosodic mimic generator also temporally aligning the prosodic mimic word with the spoken utterance, and

a processor that, if the recognized word includes the command, executes the command, and, if the recognized word includes the name to be dialed, dials a number corresponding to the name.

10. (Canceled).

11. (Previously Presented) The system of claim 9, wherein the system is disposed on a mobile telephone device.

12. (Previously Presented) The system of claim 9, further comprising a storage device including executable instructions for speech analysis and processing.

13. – 14. (Canceled)

15. (Previously Presented) The method of claim 1, wherein the command is any one of a plurality of available commands.

EVIDENCE - APPENDIX B

No evidence pursuant to §§ 1.130, 1.131, or 1.132 or entered by or relied upon by the examiner is being submitted.

RELATED PROCEEDINGS - APPENDIX C

No related proceedings are referenced in II. above, hence copies of decisions in related proceedings are not provided.